

University of Groningen

Rechtshandhaving geëvalueerd

Klein Haarhuis, C.; Ooyen-Houben, M. van; Kleemans, E.; Leeuw, F.

Published in:
Justitiële Verkenningen

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2005

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Klein Haarhuis, C., Ooyen-Houben, M. V., Kleemans, E., & Leeuw, F. (2005). Rechtshandhaving geëvalueerd: een synthese van 31 beleidsevaluaties. *Justitiële Verkenningen*, 31(8), 54.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Rechtshandhaving geëvalueerd

Een synthese van 31 beleidsevaluaties

*C. Klein Haarhuis, M. van Ooyen-Houben, E. Kleemans en F. Leeuw**

In de laatste decennia is het aantal beleidsonderzoeken, prestatie-metingen en evaluatieonderzoeken sterk in omvang toegenomen. Cijfers laten zien dat overheden en non-gouvernementele organisaties tegenwoordig vele honderden beleids- en evaluatieonderzoeken per jaar verrichten (Leeuw en Cooksey, 2004; Rist en Stame, 2005). Daarbij komen nog de 'monitoren' en 'indicatoren' die in toenemende mate worden ontwikkeld ten behoeve van de beleidsbepaling en -uitvoering (Zall Kusek en Rist, 2004).

Bij een dergelijke overvloed aan onderzoek doet zich de vraag voor of het niet mogelijk is om al deze onderzoeksresultaten te synthetiseren tot een overzichtelijk geheel. Dat dergelijke syntheses wenselijk zijn, wordt door bijna niemand betwist. Maar over hoe dit zou moeten gebeuren, lopen de meningen uiteen.

De ene stroming, die nauw verbonden is met de Campbell Collaboration Group on Crime and Justice, probeert overzicht te creëren door hoge eisen te stellen aan de kwaliteit van onderzoeksopzetten. Daarbij wordt sterk de nadruk gelegd op het belang van (quasi)-experimentele designs voor het valide en betrouwbaar meten van effecten. Of een interventie (een beleidsmaatregel) wel of niet werkt, dient in het ideale geval te worden vastgesteld door 'at random' een experimentele groep en een controlegroep samen te stellen, door vervolgens de experimentele groep te confronteren met de interventie, en door ten slotte de resultaten tussen de experimentele groep en de controlegroep met elkaar te vergelijken. De basisgedachte binnen deze stroming is dat onderzoeken die niet aan bepaalde eisen voor (quasi)-experimenteel onderzoek voldoen, gevoeglijk buiten beschouwing kunnen worden gelaten (onder anderen Farrington en Welsh, 2005).

* Dr. Carolien Klein Haarhuis, dr. Marianne van Ooyen-Houben, dr. Edward Kleemans en prof. dr. Frans Leeuw zijn allen werkzaam bij het WODC. Dit artikel is gebaseerd op het WODC-rapport *Kennis voor beleid: een synthese van 58 onderzoeken op het gebied van rechtshandhaving*, door dezelfde auteurs (Den Haag, WODC, Cahier 2005-5).

De andere stroming, waarvan Pawson (2002) één van de voortrekkers is, stelt dat bij een dergelijke benadering veel te veel bruikbare kennis buiten beschouwing wordt gelaten en dat syntheses op een bredere basis gebaseerd moeten zijn, afhankelijk van de aard van de vraagstelling én onder uitdrukkelijke bekendmaking van de aard en wijze van dataverzameling en -analyse. Pawson pleit voor een zogenoemde 'realist synthesis', omdat kennis over *hoe* interventies werken (werkzame mechanismen) minstens zo belangrijk is voor beleidsmakers als kennis over *of* interventies werken (waarbij de interventie verder een 'black box' blijft). Bovendien is deze kennis over werkzame mechanismen (binnen bepaalde contexten) veel breder toepasbaar, ook bij andere interventies en op andere beleidsterreinen.

In dit artikel willen we heel concreet aangeven hoe het maken van een dergelijke synthese in zijn werk gaat en wat de resultaten daarvan kunnen zijn. Daarbij richten wij ons op de resultaten van 31 beleidsevaluaties, die tussen januari 2002 en mei 2004 in opdracht van het Ministerie van Justitie zijn uitgevoerd op het gebied van rechtshandhaving. Met betrekking tot deze 'dwarsdoorsnede' aan evaluatieonderzoeken beantwoorden we de volgende drie vragen:

- Om wat voor typen interventies (beleidsmaatregelen) gaat het eigenlijk bij deze onderzoeken?
- Op welke manier dient de kwaliteit van deze onderzoeken te worden beoordeeld en hoe bruikbaar zijn de resultaten voor een synthese?
- Hoe kunnen de resultaten uit deze onderzoeken inhoudelijk worden gesynthetiseerd? Met andere woorden: wat hebben we inhoudelijk geleerd van twee jaar evaluatieonderzoek op het gebied van rechtshandhaving?

Rechtshandhaving: vier typen interventies

In pleidooien voor (quasi)-experimenteel onderzoek wordt vaak een vergelijking getrokken met het testen van medicijnen. Slechts weinig mensen accepteren immers een medicijn zonder dat dit medicijn uitvoerig is getest in goed opgezet experimenteel onderzoek. Waarom zouden dan voor overheidsbeleid en onderzoek naar beleidsinterventies andere normen moeten gelden?

Voordat wij een genuanceerd antwoord geven op deze vraag, is het misschien goed om eerst na te gaan hoe die beleidsinterventies er

concreet uitzien. Relatief zelden is er sprake van maatregelen die rechtstreeks zijn gericht op verandering van het nalevingsgedrag van burgers. Veel vaker blijkt er sprake te zijn van:

- interventies in de vorm van speciale projecten en nieuwe organisatievormen;
- indirecte sturing door de financiering van organisaties;
- indirecte sturing van ketens van organisaties;
- stimulering van innovatie in (ketens van) organisaties;
- invoering of wijziging van wet- en regelgeving.

In het kader van deze synthese is daarom een typologie ontwikkeld van soorten (evaluatie)onderzoek, die recht doet aan de verschillende typen beleidsinterventies en daarmee ook aan de verschillende manieren waarop dergelijke maatregelen kunnen worden geëvalueerd. Daarbij onderscheiden wij de volgende vier typen evaluaties:

- *Projectevaluaties*: Onderzoek naar de uitvoering en resultaten van concrete projecten (bijvoorbeeld het instellen van een anoniem meldpunt voor misdrijven, of een project ter bestrijding van drugsgerelateerde overlast en criminaliteit door intensivering van handhavingsinspanningen).
- *Organisatie-evaluaties*: Onderzoek naar het functioneren van specifieke organisaties (bijvoorbeeld de rol van het Openbaar Ministerie in de aanpak van milieudelicten).
- *Ketenevaluaties*: Onderzoek naar het functioneren van de gehele handhavingsketen, of van aspecten/maatregelen die deze keten aangaan (bijvoorbeeld: het functioneren van de keten 'ongebruikelijke transacties' die is geënt op de melding, verwerking, opsporing en vervolging van praktijken die kunnen duiden op bijvoorbeeld het 'witwassen' van gelden).
- *Evaluatie van wet- en regelgeving*: Onderzoek naar de werking van nieuwe wet- en regelgeving, of van wijzigingen in wet- en regelgeving (bijvoorbeeld het opheffen van het bordeelverbod en het reguleren van de prostitutiesector, of een wetswijziging ter verruiming van de toepassingsmogelijkheden van DNA-onderzoek).

In tabel 1 zijn de 31 geselecteerde evaluatieonderzoeken ingedeeld volgens de bovenstaande systematiek. De tabel geeft goed zicht op de diversiteit van de onderzochte beleidsmaatregelen. Acht onderzoeken vallen in de categorie 'projectevaluaties'. Het gaat hierbij

Tabel 1: De geselecteerde onderzoeken ingedeeld naar type (evaluatie)onderzoek

Project-evaluaties	<p><i>Overzichtelijke interventies</i></p> <ul style="list-style-type: none"> * Project Meld Misdaad Anoniem (Blauw Research, 2003). * Evaluatie team opsporing oorlogsmisdrijven; het 'NOVO-team' (Beijer e.a., 2002). * Evaluatie Kernteams, opgericht voor de uitvoering van grotere en complexere opsporingsonderzoeken (Klerks e.a., 2002). <p><i>Multimodale interventies</i></p> <ul style="list-style-type: none"> * Integratie van Financieel Rechercheren in de handhavingsketen (Faber en Van Nunen, 2002). * Drugsoverlast in Venlo; indicatoren en nulmeting (Snippe en Bieleman, 2002). * Drugsoverlast in Venlo; tussenmeting (Snippe, Bieleman en Kruize, 2004). * Indicatoren voor het monitoren van de XTC-aanpak (Snippe, Oldersma, Bieleman, 2002). * Inspannen tegen XTC: eerste meting (Snippe, Kruize en Bieleman, 2003).
Organisatie-evaluaties	<ul style="list-style-type: none"> * Openbaar Ministerie en de aanpak van milieuzaken (DHV Milieu en infrastructuur, 2002).
Keten-evaluaties	<p><i>Beschrijving huidige situatie – breed</i></p> <ul style="list-style-type: none"> * Quick scan van de handhaving in het landelijk gebied (Bos e.a., 2001). * Staat van de handhaving door gemeenten (Andersson Elffers Felix, 2002). * Handhaving afvalstoffenregeling (Frederiks e.a., 2003). <p><i>Beschrijving inefficiënties</i></p> <ul style="list-style-type: none"> * Naar redelijke doorlooptijden in de strafrechtsketen (Berenschot, 2002). <p><i>Beschrijving nieuwe vormen (eventueel pilot)</i></p> <ul style="list-style-type: none"> * Integrale handhaving van het beleid ten aanzien van openbare inrichtingen (Bekkers e.a., 2003). * Evaluatie van de keten Ongebruikelijke Transacties (Faber en Van Nunen, 2004). * Handhaving leerplichtwet via lokale partners (BBSO, 2003).
Evaluatie van wet- en regelgeving	<ul style="list-style-type: none"> * (<i>In totaal 7 studies</i>) Gevolgen van de opheffing van het bordeelverbod (6 deelonderzoeken + synthese) (Daalder, 2002). * Invoering van het Transactiebesluit Milieudelicten (Blomberg e.a., 2002). * Evaluatie van de Wet Bijzondere Opsporingsbevoegdheden (eerste fase) (Bokhorst e.a., 2002). * Evaluatie van de Wet Herziening Gerechtelijk Vooronderzoek (Verrest en Beenackers, 2002). * Handhaafbaarheid van de Meststoffenwet (Prinsen e.a., 2002). * Naleving van de Wet verontreiniging oppervlaktewateren (Prinsen en Vossen, 2002). * DNA-onderzoek in opsporing en bewijsvoering in strafzaken, nulmeting (Buiten, Dubelaar, Haesen e.a., 2003). * Doping en handel. Onderzoek naar aard en omvang van dopinghandel en ontwikkeling van indicatoren (Oldersma, Snippe en Bieleman, 2002). * Gebruik en aanschaf van cannabis door 16- en 17-jarigen (Van Ooyen-Houben en Siepermann, 2002).

concreet om zes projecten die zijn gelanceerd door Nederlandse organisaties op het gebied van rechtshandhaving. Eén onderzoek betreft een zogenoemde organisatie-evaluatie. Zeven evaluaties betreffen de handavingsketen als geheel, soms toegespitst op deelaspecten van de handhaving. Vijftien evaluaties slaan terug op de invoering of wijziging van wet- en regelgeving. Opmerkelijk is dat het bij de evaluatieonderzoeken steeds gaat om ex-post evaluaties; ex-ante evaluaties zijn niet aangetroffen.

Kwaliteit van de onderzoeken en bruikbaarheid voor een synthese

Voordat de inhoudelijke synthese heeft plaatsgevonden, hebben we eerst de kwaliteit en bruikbaarheid van de onderzoeken beoordeeld. Dit is gedaan aan de hand van de volgende vier beproefde kwaliteitscriteria.

In de eerste plaats is de onderzoekskwaliteit beoordeeld aan de hand van het meer algemene criterium van 'descriptieve validiteit' (Wittebrood en Van Beem, 2004). Gefocust is op de beschrijving van de onderzoeksopzet, de meetmethode en van de eventuele selectiemethode. Ook is nagegaan of meerdere informatiebronnen zijn gebruikt, omdat dit de betrouwbaarheid van de onderzoeksbevindingen verhoogt (Yin, 1994).

In de tweede plaats hebben we beoordeeld of in een onderzoek is nagegaan of en in welke mate een beleid daadwerkelijk is geïmplementeerd. Want beleid dat nooit is uitgevoerd, kan ook niet zinvol worden geëvalueerd. Eerder onderzoek laat zien dat de programma-integriteit, de daadwerkelijke implementatie van een beleid conform plan, vaak te wensen overlaat (Lösel, 1995). Ook een recent rapport van de Algemene Rekenkamer (2003) toont aan dat er veel meer beleidsmaatregelen worden gemaakt dan er daadwerkelijk worden uitgevoerd.

In de derde plaats is geanalyseerd of in de onderzoeken aandacht wordt besteed aan de *veronderstellingen* die ten grondslag liggen aan beleidsinterventies (onder anderen Leeuw, 2003; Pawson, 2002). Bij het ontwerpen en ontwikkelen van beleid worden reacties op interventies verwacht van de doelgroep (burgers, [potentiële] criminelen, politieagenten enzovoort). Beleidsmakers veronderstellen verder ook gedrag van derden (bijvoorbeeld 'buren', advocaten) en ze veronderstellen effecten. Van evaluatoren wordt verwacht dat

zij de assumpties over hoe beleid zou moeten werken, articuleren (de '*programmatheorie*'), en deze vervolgens testen. *Hoe* (via welke mechanismen) wordt beleid verondersteld te werken, en tot welke resultaten zou dit moeten leiden? Pawson en Tilley (1997) houden een pleidooi om beleidsinterventies én evaluatieonderzoek te structureren aan de hand van Context- Mechanismen- Outcome ('C-M-O')-configuraties. Beleid(interventies) worden verondersteld zich te onttollen in bepaalde contexten. Denk aan hoe de rechtshandhaving is georganiseerd, of er sprake is van centrale én mede-overheden of juist niet, hoe omvangrijk de misdaad is enzovoort. Deze contexten hebben op zichzelf genomen geen 'wervingskracht'. Dat hebben de gedragsmechanismen die in de verschillende interventies besloten liggen. Denk aan interventies die de werking van (financiële) prikkels veronderstellen, interventies die juist voortborduren op vertrouwen tussen organisaties en medewerkers, interventies die regels en sancties centraal stellen, interventies die een beroep doen op de behoefte van mensen cognitieve dissonantie te voorkomen of te reduceren enzovoort (voor een categorisering zie Hedström en Swedberg, 1998). Ten slotte wordt verondersteld dat de combinatie van Contexten & Mechanismen bepaalde resultaten ('outcome') tot gevolg zal hebben. Denk aan een betere opsporing of het terugdringen van misdaad. Het belang van een dergelijke benadering is tweërlei. Enerzijds leidt het tot betere onderzoeksopzetten, omdat ook de werkzame mechanismen en de tussenstappen tussen interventie en effect worden onderzocht. Anderzijds is kennis over deze mechanismen veel breder toepasbaar, ook bij andere interventies en op andere beleidsterreinen.

In de vierde en laatste plaats is bij evaluatieonderzoek de vraag van belang naar interne validiteit: of de beoogde resultaten van de onderzochte beleidsinterventie door die interventie tot stand zijn gekomen, of door iets anders, met andere woorden, of een bepaalde beleidsinterventie de beoogde resultaten heeft veroorzaakt (Cook en Campbell, 1979; Sherman e.a., 2002). Dit criterium kan worden geplaatst in de benadering van de Campbell Collaboration Group on Crime and Justice, die we hierboven introduceerden. Hoe zekerder dit kan worden vastgesteld, des te hoger de interne validiteit, en des te meer gesproken kan worden van effecten in strikte zin. Voor de beoordeling van deze interne validiteit van de onderzoeken hebben we gebruikgemaakt van de Maryland Scientific Methods Scale (MSMS), een vijf puntsschaal (Sherman e.a., 2002; Farrington e.a., 2002) (zie tabel 2).

Tabel 2: Vijf niveaus van interne validiteit conform de Maryland Scientific Methods Scale

Niveau 1	Samenhang tussen een beleidsinterventie en uitkomstvariabele na invoering van de beleidsinterventie
Niveau 2	Score op de uitkomstvariabele voor en na invoering van de beleidsinterventie, zonder (vergelijkbare) controlegroep
Niveau 3	Score op de uitkomstvariabele voor en na invoering van de beleidsinterventie, gemeten in een experimentele en vergelijkbare controlegroep (quasi-experimenteel design)
Niveau 4	Niveau 4: Score op de uitkomstvariabele voor en na invoering van de beleidsinterventie gemeten in een quasi-experimentele setting, met controle voor andere variabelen die de uitkomstvariabele beïnvloeden
Niveau 5	Score op de uitkomstvariabele voor en na invoering van de beleidsinterventie, waarbij de beleidsmaatregel 'at random' is toebedeeld aan een experimentele en een controlegroep (experimenteel design).

Elk niveau hoger op deze schaal betekent een grotere mate van zekerheid waarmee resultaten aan beleid kunnen worden toegeschreven (interne validiteit). Op niveau 1 is die zekerheid nihil. In feite wordt alleen gekeken naar de uitkomstvariabele (bijvoorbeeld mate van jeugdcriminaliteit) op een bepaald punt in de tijd, zonder een vergelijking te trekken met de periode vóór invoering van de beleidsinterventie, of met situaties waarin deze niet is toegepast. Op niveau 2 wordt de uitkomstvariabele ná invoering van de beleidsinterventie wél vergeleken met de situatie ervóór. Maar ook dan kan niet meer worden geclaimd dan een eventuele verandering in de uitkomstvariabele *ten tijde van* het beleid, die dus ook aan iets anders zou kunnen worden toegeschreven dan aan de beleidsinterventie. Op niveau 3 spreken we van een quasi-experimenteel design. De uitkomstvariabele wordt niet alleen vergeleken met de situatie vóór invoering van de beleidsinterventie, maar ook met een 'controlegroep' die niet met het beleid is geconfronteerd. Volgens Cook en Campbell (1979) vormt dit het minimumniveau waarmee op verantwoorde wijze conclusies kunnen worden getrokken over de effectiviteit van een interventie. Op niveau 4 wordt daarbij nog eens gecontroleerd voor specifieke andere factoren, die volgens de onderzoeker de uitkomstvariabele mogelijk beïnvloeden (bij criminaliteit valt te denken aan sociaal-psychologische kenmerken, opvoeding, opleiding). Farrington en Welsh (2005) hanteren het criterium van

experimenteel design (niveau 5) voor het opnemen van studies in hun onderzoekssynthese. Met het 'at random' toebedelen van de beleidsinterventie aan experimentele en controlegroep worden automatisch alle mogelijke externe invloeden uitgeschakeld, óók van factoren die de onderzoeker niet bekend zijn.

Met de bovengenoemde vier criteria beoordelen wij de onderzoeken breed op hun merites. Er ontstaat zo meer aandacht voor de inhoud en de beschrijvende kwaliteit van de onderzoeken dan wanneer we alléén de MSMS als maatstaf zouden gebruiken.

De kwaliteit van de 31 onderzoeken

Hoe goed zijn de onderzoeken als we ze beoordelen aan de hand van de vier criteria? Het blijkt dat de onderzoeken op het ene criterium beter scoren dan op het andere, maar al met al komen ze er redelijk uit. Appendix 1 illustreert dit aan de hand van de projectevaluaties. De meeste onderzoeken maken gebruik van gegevens uit meerdere informatiebronnen. Door verschillende gegevens te combineren, geven de onderzoekers vaak een genuanceerd beeld van hun onderwerp. Dit is een sterk punt van de onderzoeken, het maakt hun resultaten stevig en valide.

Ook brengen alle onderzoeken in beeld of de onderzochte beleidsmaatregel is geïmplementeerd. Hierdoor kan met enige zekerheid aangenomen worden dat een gevonden resultaat verband houdt met de beleidsinterventie. Dit helpt ons om de causaliteitsvraag beter te kunnen beantwoorden. Ook op dit punt zitten de onderzoeken gedegen in elkaar.

Verder vonden we dat in ruim éénderde van de evaluatieonderzoeken expliciet is beschreven langs welke weg een nieuwe beleidsinterventie tot de beoogde resultaten zou moeten leiden. Met andere woorden, deze onderzoeken bevatten een rudimentaire weergave van de 'programmatheorie'. De onderzoeken geven vaak heldere analyses van de werking van het beleid in de praktijk. Maar een expliciete analyse waarmee 'veronderstelde' met 'werkzame' mechanismen worden vergeleken, die dieper gaat dan een beschrijving van beoogde interventies en beoogde doelen, hebben we nauwelijks aangetroffen. Dit punt is minder sterk.

Afgemeten aan de MSMS-schaal bevinden de meeste evaluaties zich op niveau 1 of 2. Meestal verantwoorden de onderzoekers niet waarom ze geen controleconditie (nulmeting of controlegroep)

hebben gebruikt. Het is duidelijk dat de onderzoeken op het punt van interne validiteit niet sterk zijn. Het is wel zo dat de MSMS soms niet van toepassing is omdat een effectmeting niet is beoogd. Ook is het soms niet mogelijk om een geschikte controleconditie te vinden, bijvoorbeeld omdat een maatregel landelijk is ingevoerd. Maar hoe het ook zij, de score op de MSMS is laag, en we kunnen ons afvragen of dit in een aantal onderzoeken niet beter had gekund.

Waarom synthese?

Tussen de 31 evaluatieonderzoeken zijn geen evaluatiestudies van niveau 3 of hoger gevonden. Menigeen zal daarom stellen dat het synthetiseren van bevindingen niet raadzaam is. Wij willen dit bestrijden. Waarom?

In de eerste plaats komt de descriptieve validiteit van de meeste studies tegemoet aan algemene evaluatiestandaarden (meerdere onderzoeksmethoden, accuraatheid van bevindingen). In de tweede plaats wordt in vrijwel elk van de 31 studies aandacht besteed aan het implementatieproces. In de derde plaats wordt in veel studies aandacht besteed aan onderliggende werkingsmechanismen; zodra het beleid is geïmplementeerd. Hiermee valt enig zicht te krijgen op de 'black box' van beleidsinterventies: wat is het in een beleid dat de uitkomst veroorzaakt, hoe komt deze tot stand, voor wie, en onder welke condities? Als het doel van een synthese van evaluatieonderzoeken is erachter te komen of interventies daadwerkelijk zijn ingevoerd of als de taak is achter de werkingsmechanismen te komen, dan zou het onverstandig zijn om bevindingen beneden niveau 3 van de MSMS niet te synthetiseren (vergelijk Pawson, 2002). Dit wordt nog eens onderstreept door het toenemende belang dat wordt gehecht aan een 'what works-' of kennisgestuurde aanpak, vooral ook op het gebied van criminaliteit en rechtshandhaving. Een reconstructie van beleids-theorieën en een blik op werkzame mechanismen verhoogt niet alleen de kwaliteit van evaluaties, maar ook de bruikbaarheid van evaluatieonderzoek voor beleid.

Inhoudelijke synthese

Hoe gaat nu een inhoudelijke synthese van de onderzoeksresultaten in zijn werk en wat levert dit uiteindelijk op aan kennis? Voor elk evaluatieonderzoek hebben we geïnventariseerd:

- de inhoud van het beleid (respectievelijk de interventies), zoals afgeleid uit de evaluatierapporten;
- de mechanismen op basis waarvan het beleid wordt verondersteld te werken, en de condities waaronder het beleid tot resultaten zou moeten leiden;
- de feitelijke werking van het geëvalueerde beleid, al dan niet gekoppeld aan voornoemde mechanismen;
- contextuele condities waaronder het beleidsproces al dan niet effectief blijkt te zijn;
- het feitelijk behalen van de met het beleid beoogde doeleinden: welke resultaten zijn bereikt?

Appendix 2 illustreert dit aan de hand van de projectevaluaties. Vervolgens zijn de bevindingen uit de onderzoeken inhoudelijk gesynthetiseerd, dat wil zeggen samengebracht op een niveau dat de individuele onderzoeken overstijgt, vorsend naar algemene lijnen en rode draden.

Wat leert deze inhoudelijke synthese van onderzoeken ons nu over de rechtshandhaving in Nederland? Wat kunnen we leren over de uitvoering van projecten, over maatregelen die de handhavingsketen willen verbeteren en over de invoering van nieuwe wetten en regels?¹

Projecten op het gebied van rechtshandhaving bestaan vaak uit verhoogde input van mankracht, expertise en middelen, en uit intensivering van samenwerking. Beleidsmakers vinden het belangrijk dat projecten goed aansluiten bij de bestaande ketenorganisatie. Daarom worden ketenpartners zoals politie en Openbaar Ministerie actief bij een project betrokken. Een aantal projecten behelzen extra sturing of overleg. Door dit alles zou betere informatie beschikbaar moeten komen en zouden handhavingsinstrumenten beter gebruikt moeten worden. Dit moet dan leiden tot meer opsporingsonderzoeken, aanhoudingen, veroordelingen, en uiteindelijk tot een verhoogde pakkans en minder criminaliteit.

Wat zien we in de praktijk? De verhoogde input en de betere samenwerking blijken goed gerealiseerd te worden. De informatiepositie van betrokken organisaties wordt vaak versterkt. Maar we vonden een aantal systematische knelpunten. Er treden vaak problemen op in de aansturing, wat leidde tot fragmentatie en inefficiënties.

¹ Op de categorie organisatie-evaluaties gaan we niet in omdat er maar één onderzoek in de steekproef zit.

Ook was er in enkele projecten gebrek aan scholing en expertise van medewerkers of verloop onder personeel. Organisaties blijken vaak terughoudend met het oppakken van de aan nieuwe projecten verbonden activiteiten. De projectkennis wordt dan niet omgezet in daden. Het nieuwe dat het project heeft opgeleverd raakt moeilijk ingebed in de bestaande werkprocessen. Dit kan duiden op een variant van het bekende 'kennis-attitude-gedragsprobleem': via communicatie opgedane projectkennis in een organisatie hoeft nog niet te betekenen dat die kennis ook in daden wordt omgezet.

In de meeste projecten komen meer opsporingsonderzoeken, aanhoudingen en andere 'producten' tot stand. De einddoelen zijn vaak een zaak van lange adem. De evaluatieonderzoeken van projecten die wij analyseerden, laten op het punt van de einddoelen nog nauwelijks resultaten zien. Soms is het nog te vroeg daarvoor.

Interventies in de keten beslaan meestal de hele keten van preventie, opsporing en vervolging tot en met sanctietoepassing. Het gaat vaak om de invoering van nieuwe regulering, handhavingsinstrumenten en -methoden en om verbetering van de communicatie en samenwerking tussen ketenpartners. Dit zou moeten leiden tot betere opsporings- en vervolgingsresultaten en uiteindelijk tot minder criminaliteit.

Wat zeggen de onderzoeken hierover? De onderzoeken stellen (onder meer) vast dat er systematische knelpunten zijn die bij verschillende vormen van ketenhandhaving telkens terugkeren. Dit zijn met name:

- gebrek aan aansturing van de organisatie; te weinig integrale besturing;
- gebrek aan coördinatie tussen ketenpartners en met externe partners;
- gebrek aan gelijkvormigheid in beleid tussen ketenpartners (verschillende stijlen, werkwijzen, beleidscycli en prioriteiten); deelbewerkingen in de keten zijn te weinig in lijn met elkaar;
- gebrek aan organisatiecapaciteit;
- gebrek aan deskundigheid (opleidingsniveau);
- teveel focus op misdaadbestrijding, ten koste van preventie en kennisverwerving.

Door deze knelpunten werken de mechanismen in de praktijk niet zoals dat vooraf door het beleid werd beoogd.

Wat zijn de resultaten van keteninterventies? Het blijkt dat de organisatie en de uitvoering van de handhaving in een aantal gevallen onder de norm worden bevonden. Op gemeentelijk niveau wordt

bijvoorbeeld veel gedoogd. De ketenstructuur vertoont over het algemeen veel inefficiënties. Daarnaast lijken zowel het beleid als de evaluaties een blinde vlek te vertonen. Rechtshandhavingsbeleid draait uiteindelijk om het nalevingsgedrag van burgers en rechtspersonen, maar veel interventies en ook veel evaluatieonderzoek is daar niet rechtstreeks op gericht. De focus betreft voornamelijk de uitwerking van beleidsingrepen op de handhavingsketen, terwijl de uiteindelijke effecten hiervan op het nalevingsgedrag veelal buiten beschouwing blijven.

Wat kunnen we, ten slotte, leren uit de evaluaties van wet- en regelgeving en wijzigingen hierin? Hoewel de veelvormigheid aan geëvalueerde wet- en regelgeving groot is, komen er enkele algemene lijnen naar voren.

Meestal blijkt het bij nieuwe wet- en regelgeving te gaan om uitbreiding van het handhavingsinstrumentarium, soms gaat het om stroomlijning van bevoegdheden. Beleidsmakers verwachten dat dit zal leiden tot ruimere toepassing van opsporings- en handhavingsmogelijkheden en tot verhoogde efficiëntie van de handhaving.

Hogere ophelderingspercentages en uiteindelijk minder crimineel gedrag moeten uiteindelijk het gevolg zijn.

Het daadwerkelijke gebruik van een aantal verruimde handhavingsmogelijkheden is in de praktijk gering en vertoont knelpunten. Dit kan te maken hebben met het feit dat de evaluatieonderzoeken kort na invoering van de wetten of wetswijzigingen plaatsvonden. Betrokkenen in de praktijk zelf zijn positief over de doelbereiking van de wet- en regelgeving; het handhavingsinstrumentarium is completer geworden en bepaalde vormen van criminaliteit zijn beter te bestrijden.

Besluit

Onze exercitie heeft laten zien dat het evaluatieonderzoek op het gebied van rechtshandhaving goede descriptieve kwaliteit heeft. Ook geeft het zicht op implementatie en resultaten van beleid. Het vormt dan ook een geschikte basis om kennis voor beleid uit op te doen die het niveau van het individuele onderzoek overstijgt. Het belang van het verhogen van de kwaliteit van evaluatieonderzoek staat niettemin buiten kijf. Het gedachtegoed van de Maryland-standards over het waarborgen van de interne validiteit aan de hand

van (quasi-)experimentele designs vormt daarbij een belangrijke leidraad.

Dit neemt niet weg dat deze concrete exercitie ons leert dat er meer onder de zon is dan Maryland-standards. In de eerste plaats hebben we daarom gewezen op het belang van onderzoek naar de feitelijke implementatie van maatregelen. Want beleid dat nooit begonnen is, kan ook niet zinvol worden geëvalueerd. Bovendien kan bijsturing van beleid ook al in de implementatiefase plaatsvinden.

In de tweede plaats hebben wij gewezen op het belang van onderzoek naar de werkzame mechanismen, aan de hand waarvan men kan laten zien *waarom* beleid wel of niet werkt. De zogenoemde 'black box' van beleidsinterventies en hun uitwerking dient door beleidsmakers en onderzoekers te worden geopend. Een reconstructie van beleidstheorieën en een blik op werkzame mechanismen verhoogt niet alleen de kwaliteit van evaluaties, maar ook de bruikbaarheid van evaluatieonderzoek voor beleid. Kennis over deze mechanismen laat namelijk zien welke elementen van beleid essentieel zijn. Daarbij maakt het in feite niet uit of deze kennis berust op 'succesvol' of 'mislukt' beleid.

Voorals wanneer evaluatieonderzoek inzicht verschaft in *hoe* beleid werkt *én of* het effectief is, biedt het kennis over preventie en bestrijding van criminaliteit waarop beleidsmakers beslissingen kunnen baseren.

Literatuur

Algemene Rekenkamer

Tussen beleid en uitvoering; lessen uit recent onderzoek van de Algemene Rekenkamer

Den Haag, Kamerstukken II, 2002-2003, 28 831 nrs. 1-2, 2003

Cook, T.D., T.D. Campbell

Quasi-experimentation; design and analysis issues in field settings

Chicago, Rand McNally/ Boston, Houghton-Mifflin, 1979

Farrington, D.P., D.C.

Gottfredson e.a.

The Maryland Scientific Methods Scale

In: Sherman, L. W., Farrington e.a. (red.), *Evidence-based crime prevention*, London, Routledge, 2002, p. 13-21

Farrington, D.P., B.C. Welsh

Randomized experiments in criminology; what have we learned in the last two decades?

- Journal of experimental criminology, nr. 1, 2005, p. 9-38
- Hedström, P., R. Swedberg**
Social mechanisms; an introductory essay
 In: Hedström, P., R. Swedberg (red.), *Social mechanisms; an analytical approach to social theory*, New York, Cambridge University Press, 1998
- Klein Haarhuis, C.M. , M. van Ooyen, E.R. Kleemans en F.L. Leeuw**
Kennis voor beleid; een synthese van 58 onderzoeken op het gebied van rechtshandhaving
 Den Haag, WODC, Cahier 2005-5
- Leeuw, F.L.**
Reconstructing program theories; methods available and problems to be solved
 American journal of evaluation, 24e jrg., nr. 1, 2003, p. 5-20
- Leeuw, F.L., L. Cooksey**
Evaluating the performance of development agencies; the role of meta-evaluations
 In: World Bank, *Proceedings of the Fifth Conference on Evaluations*, Washington, World Bank, 2004
- Lösel, F.**
The efficacy of correctional treatment; a review and synthesis of meta-evaluations
 In: McGuire, J. (red.), *What works; reducing re-offending*, Chichester, UK, John Wiley & Sons, 1995, p. 79-114
- Pawson, R., N. Tilley**
Realistic evaluation
 Londen/Thousand Oaks/New Delhi, Sage Publications, 1997
- Pawson, R.**
Evidence-based policy; the promise of 'realist synthesis'
 Evaluation, 8e jrg., nr. 3, 2002, p. 340-358
- Rist, R.C., N. Stame**
From studies to streams
 New Brunswick, Transaction Publishers, 2004
- Sherman, L.W., D.P. Farrington e.a. (red.)**
Evidence-based crime prevention
 London, Routledge, 2002
- Wittebrood, K., M. van Beem**
Sociale veiligheid vergroten door gelegenheidsbeperking; wat werkt en wat werkt niet?
 Den Haag, Sociaal Cultureel Planbureau i.o.v. Raad voor Maatschappelijke Ontwikkeling, 2004, Werkdocument 2
- Yin, R.K.**
Case study research, design and methods
 Newbury Park, Sage Publications, 1994
- Zall Kusek, J., R.C. Rist**
Ten steps to a results-based monitoring and evaluation system; a handbook for development practitioners
 Washington, The World Bank, 2004

Appendix 1: De kwaliteit van projectevaluatieonderzoek

Interventie (X) en beoogd effect (Y)	Opzet evaluatie (descriptieve validiteit)	Beschrijving van de implementatie?	Reconstructie en evaluatie van beleidsveronderstellingen (mechanismen, resultaten)	MSMS (1-5) Toelichting
Project M. Instelling van een anoniem meldpunt misdrijven moet leiden tot meer opsporingen en meer vervolgingen.	<ul style="list-style-type: none"> Inventarisatie en proces-beschrijving van binnengekomen meldingen in 5 politieregio's gedurende de eerste 8 maanden van het project. Media-analyse van ongeveer 300 artikelen Steekproef onder de bevolking voor interviews (voor en na) t.a.v. attitudes en perceptie van M. 	Ja.	Beschrijving van de werking met strakke proceslijn (van communicatie tot meldingen, tot en met verwerking en gebruik van de meldingen). Hierdoor kunnen effecten met meer zekerheid worden toegeschreven aan het beleid.	Niveau 2 Nulmeting en tussenmeting beschikbaar. Geen controlecondities.
NOVO-team voor oorlogsmisdrijven moet leiden tot opsporing, intermediair en bewijsvergaring.	<p>Tussenevaluatie van effecten, aan de hand van documentstudie en interviews.</p> <p>Bestudering van de condities waaronder NOVO heeft gefunctioneerd:</p> <ul style="list-style-type: none"> Contextanalyse Interne analyse Triangulatie en rapportage. 	Ja.	Beschrijving van de werkingsmechanismen, met confrontatie tussen geformuleerde en uitgevoerde taken.	Niveau 2 Nulmeting en tussenmeting beschikbaar. Geen controlecondities.
Kernteams: gespecialiseerde opsporings-teams dienen complexer en grootschaliger zaken aan te pakken en op te lossen.	<p>Beschrijvende studie van de 'toegevoegde waarde' van de Kernteams.</p> <ul style="list-style-type: none"> Deskonderzoek Geen dossieronderzoek (is verantwoord) Interviews in alle 7 Kernteams onder functionarissen (slechts 8 per team) en enquêtes onder alle medewerkers. 	Ja.	Geen reflectie op centrale mechanismen van beleid, maar focus op implementatieproces en organisatie.	Niveau 1 Nulmeting ontbreekt. Het ontbreken van een nulmeting is niet expliciet verantwoord. Geen controlecondities. <i>De facto</i> dus een doelbereikings-onderzoek. Wel wordt meestal goed verantwoord waarom informatie gebrekkig is.

Financieel Rechercheren: opleiding en communicatie tussen ketenpartners (politie, O.M.) moet leiden tot acceptatie, integratie en implementatie van financieel onderzoek. Dit ter bestrijding van georganiseerde criminaliteit en fraude.	Studie naar de mate van implementatie na 5 jaar, met als hoofdindicator: de integratie van FR in de handhavingsketen. * Kwalitatief: Interviews (n=155), enquêtes (n=833), trefwoorden-brainstorm, documentstudie * Kwantitatief: Compas-gegevens, IFT-registraties, BOOM, CJIB. Voor evaluatie van het onderwonderdeel is een steekproef getrokken. Niet representatief voor politieregio's en parketten.	Ja.	De feitelijke werkingsmechanismen worden niet sterk teruggekoppeld naar veronderstelde mechanismen, maar staan verdeeld over de deelrapporten.	Niveau 1 Nulmeting ontbreekt. Geen controlecondities. Omdat FR het enige project over dit onderwerp is, zijn sommige resultaten automatisch met enige zekerheid aan het project FR toe te schrijven.
Project 'Hektor' moet drugsoverlast in Venlo terugdringen met behulp van een multimodale aanpak.	* Verzameling van registratiegegevens politie, belastingdienst, O.M. * Interviews met ondernemers * Peiling drugsoverlast onder bevolking * Observaties door stadswachten aan de hand van schema.	Ja, bij eindmeting.	Beschrijving van de veronderstelde aanpak en uitkomsten, indicatoren op basis hiervan vastgesteld. Er vindt terugkoppeling naar veronderstellingen plaats.	Niveau 2 Opzet met nulmeting, tussenmeting en eindmeting. Geen controlecondities.
'Samenspannen tegen XTC' moet de XTC-productie en -handel in en vanuit Nederland terugdringen.	* Eerst zijn indicatoren geïnventariseerd voor meten van input, samenwerking en resultaten * De indicatoren zijn gescreend op validiteit en beschikbaarheid * 59 indicatoren gekozen voor de eerste meting gebaseerd op plannen van de aanpak XTC (beoogde input, processen, output en 'outcome') * Gegevens uit registratiesystemen van betrokken actoren * Schriftelijke vragenlijst over samenwerking.	Ja, bij tussenmeting.	Veronderstellingen worden beschreven en indicatoren op basis hiervan vastgesteld. Bevindingen worden teruggekoppeld naar veronderstellingen. De studie laat zien wat er van de geplande aanpak tot stand is gekomen.	Niveau 2 Opzet met nulmeting, tussenmeting en eindmeting. Geen controlecondities.

Appendix 2: De inhoud van projectevaluatieonderzoek: interventie(s) en beoogde effecten, veronderstelde mechanismen, bevindingen m.b.t. mechanismen en andere factoren, resultaten

Projecten	Interventie(s) en beoogde effecten	Veronderstelde mechanismen	Feitelijke werking van mechanismen; overige factoren	Resultaten
Project 'M.' (Meld Misdaad Anoniem)	Instelling van een anoniem meldpunt misdrijven moet leiden tot meer opsporingen en meer vervolgingen.	<ul style="list-style-type: none"> Communicatie 'M' Bekendheid bij burger Attitudes Intentie tot gedrag Feitelijk gedrag bij burgers Frontoffice: meldpunt Backoffice: verwerken van bruikbare meldingen Resultaat: opsporing, vervolging. 	<ul style="list-style-type: none"> Knelpunt is de beslissingsbevoegdheid van 'backoffices' Van welke <i>typen</i> delicten meldingen zijn gewenst, is beter te sturen. 	<ul style="list-style-type: none"> Frontoffice, 2000 gesprekken Burgerattitudes positief 6 tot 9% onderzoek 1-2% resultaat in de opsporing Geen langetermijneffecten beschikbaar.
Evaluatie NOVO-team oorlogsmisdrijven	Het NOVO-team is opgericht ter opsporing van oorlogsmisdrijven, intermediering en bewijsvergaring.	<ul style="list-style-type: none"> Aangifte Bewijsvergaring Verhoor Institutionele inbedding: aansluiting bij ketenpartners Rapportage. 	<ul style="list-style-type: none"> Aard van de zaken is problematisch (moeilijk bewijsbaar) Goede kwaliteit NOVO-team Lage prioriteit bij ketenpartners. 	Het NOVO-team heeft geen nieuwe opsporingen gegenereerd.
Evaluatie Kernteams	Instelling van gespecialiseerde opsporingsteams moet leiden tot meer, complexere, en grootschalige opsporingsonderzoeken.	<ul style="list-style-type: none"> Expertise en mankracht verenigen in speciale teams. 	<ul style="list-style-type: none"> Veel verschillen tussen kernteams Gebrek aan landelijke strategische aansturing Positie 'uit de wind' blijkt gunstig voor complexiteit, expertise, samenhang en continuïteit Medewerkers blijken te licht opgeleid Kernteams blijken onontbeerlijk als communicatieknooppunt. 	Geen eenduidige conclusie mogelijk over de toegevoegde waarde van de kernteams.

Financieel Rechercheren	Opleiding en communicatie tussen ketenpartners (politie, O.M.) moet leiden tot acceptatie, integratie en implementatie van financieel rechercheren. Ter bestrijding van georganiseerde criminaliteit en fraude.	<i>Sub-doelstellingen:</i> Bestuurlijk-organisatorische adviezen * Communicatieverbetering * Opleidingen * Methodenontwikkeling * Aanpassing wet- en regelgeving * Ontwikkeling van preventieve maatregelen.	<ul style="list-style-type: none"> * Doelstelling niet helder en te ambitieus * Beleving van delicten als niet ernstig * Te gering zaaksanbod * Gebrek aan aansturing. 	<ul style="list-style-type: none"> * Financieel rechercheren is op de agenda gezet bij politie en Openbaar Ministerie * Wel geïntegreerd onder specialisten, maar niet in de gemiddelde politieregio.
Project 'Hektor' tegen drugs-overlast in Venlo	Drugsoverlast terugdringen met behulp van een multimodale aanpak.	<ul style="list-style-type: none"> * Intensivering van de handhaving: verhoogde input van mensen en middelen (hogere pakkans) * Aanpak van drugsgerelateerd vastgoed * Aanpassing coffeeshopbeleid. 	<ul style="list-style-type: none"> * Inspanningen en 'productie' (bijv. aanhoudingen) van betrokken diensten zijn toegenomen * Aanpassing coffeeshopbeleid nog niet gerealiseerd ten tijde van tussenmeting. 	<ul style="list-style-type: none"> * Overlast en criminaliteit bij tussenmeting nog niet verminderd * Geregistreerde drugsbanden, -dealers en -runners afgenomen * Groter gevoel van veiligheid en meer economische activiteit.
Evaluatie Nota 'Samen-spinnen tegen XTC'	Reductie van productie en handel XTC in en vanuit Nederland.	<ul style="list-style-type: none"> * Intensivering van de aanpak van XTC: verhoogde input van mensen en middelen (hogere pakkans) * Aanpak van precursoren en hulpmiddelen, productielocaties en handel. 	<p>Eerste resultaten overwegend in de beoogde richting:</p> <ul style="list-style-type: none"> * Inspanningen en 'productie' (bijv. ontmantelingen) door betrokken diensten toegenomen * De informatiepositie van Douane, FIOD-ECD is versterkt * De samenwerking van opsporings- en controlediensten is geïntensiveerd. 	<ul style="list-style-type: none"> * Meer inbeslagnames in het eerste jaar van implementatie * Meer productieplaatsen ontmanteld * Minder milieudumping * Vooralsnog geen nadere resultaten (alleen eerste meting verricht).